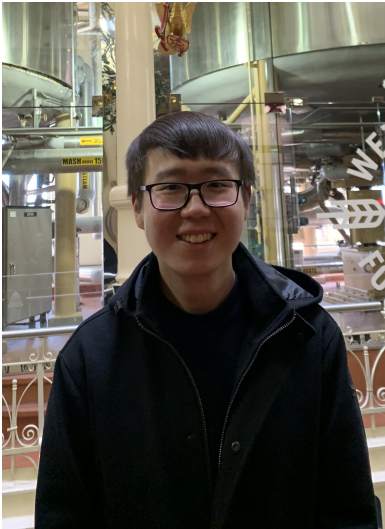




Special Colloquium



Estimating the Background Component in a Two-Component Mixture Models

He Jiang

University of California, San Diego

Abstract: Cluster analysis is broadly seen as one of the most important tasks in (unsupervised) data analysis with a broad range of applications. It focuses on the task of grouping a set of objects into clusters, such that objects in the same cluster are more similar to each other than to those in different clusters. This grouping could either be hard, i.e. one object can only belong to one cluster, or soft, i.e. one object can belong to multiple mixture components with probability summing to 1. In this talk, we will focus on the later situation and consider a two-component mixture model, where one component plays the role of the background component. This model is commonly seen in multiple testing and contamination models, and has many applications in areas such as astronomy, biology, and social sciences. Extending on the current literature, we consider three mild assumptions on the background distribution: when it is symmetric; when it is monotonic; and when it is log-concave. In each setting, we derive estimators for the background component, and provide relevant characteristics of these estimators. We also include some possible future projects.

Keywords: Clustering; two-component mixture models; background distribution; multiple testing.

Tuesday, February 8, 12:05 – 12:50 pm

For more info visit the [department website for the colloquium](#)